

АЛГОРИТМ ПОСТРОЕНИЯ ВИРТУАЛЬНОГО АНАЛИЗАТОРА НА ОСНОВЕ РАСШИРЕНИЯ ОБУЧАЮЩЕЙ ВЫБОРКИ

Д.В. Штакин

Институт автоматики и процессов управления ДВО РАН
Россия, 690041, Владивосток, ул. Радио, 5
E-mail: shtakin@iacp.dvo.ru

А.Ю. Торгашов

Институт автоматики и процессов управления ДВО РАН
Россия, 690041, Владивосток, ул. Радио, 5
E-mail: torgashov@iacp.dvo.ru

Ключевые слова: виртуальный анализатор, малая выборка, расширение выборки, нейронная сеть, ректификация.

Аннотация: Решается задача повышения точности виртуальных анализаторов показателей качества продуктов ректификационных технологических процессов в условиях малой обучающей выборки. Недостаточный объем обучающей выборки как правило приводит к переобучению разрабатываемых математических моделей в составе виртуальных анализаторов. Предложен алгоритм расширения малой выборки данных по показателю вязкости среднего дистиллята за счет использования более доступных данных лабораторного контроля по его фракционному составу, что, в итоге, позволило достичь более высокой точности виртуального анализатора для целей непрерывного контроля качества конечной продукции технологического процесса с использованием систем усовершенствованного управления.

1. Введение

Виртуальные анализаторы (ВА) получили широкое распространение на нефтеперерабатывающих производствах как инструмент для повышения качества конечной продукции с минимальными затратами при заданных ограничениях в составе усовершенствованных систем управления. С использованием ВА повышается эффективность управления технологическим объектом, т.к. обычно лабораторные анализы проводятся 1-2 раза в сутки, что недостаточно для обеспечения высокой оперативности управления.

В состав ВА входят математические модели для оценивания трудноизмеримых целевых показателей на основе легкодоступных данных технологического мониторинга (температурный профиль, давления, расходы потоков). Как правило в качестве ВА на производстве используются линейные модели, строящиеся по данным технологического мониторинга и лабораторного анализа. Однако такие показатели качества как вязкость выходного продукта, как правило, анализируются лабораторией значительно реже, чем его фракционный состав. Это приводит к дефициту данных для построения виртуального анализатора обеспечивающего достаточную точность оценок показателей качества, что является распространенной проблемой [1].

Задача заключается в повышении точности ВА по вязкости среднего дистиллята при 40°C в условиях малой обучающей выборки.

3. Предлагаемый алгоритм расширения обучающей выборки

Для преодоления проблемы недостаточного объема обучающей выборки предлагается алгоритм ее расширения на основе данных более частоизмеряемых показателей качества (фракционный состав) выходного продукта.

Предлагаемый алгоритм заключается в следующем:

- **шаг 1:** корреляционный анализ данных лабораторного анализа редкоизмеряемого и частоизмеряемых показателей качества для определения входных переменных промежуточной модели;
- **шаг 2:** построение промежуточной (вспомогательной) модели редкоизмеряемого показателя качества на основе выбранных частоизмеряемых показателей качества;
- **шаг 3:** получение дополнительного сегмента данных (ДСД) обучающей выборки при помощи промежуточной модели на основе данных частоизмеряемых показателей качества;
- **шаг 4:** построение виртуального анализатора на основе выборки, содержащей ДСД.

В таблице 1 представлены результаты проведенного корреляционного анализа лабораторных данных. Используемые лабораторные данные о фракционном составе содержат 318 наблюдений по каждому из следующих показателей: температура начала кипения (ТНК), температура выкипания 10% (ФС 10%), температура выкипания 50% (ФС 50%), температура выкипания 90% (ФС 90%), температура выкипания 95% (ФС 95%). Лабораторные данные по вязкости при 40°C (Вязкость) содержат 29 наблюдений. В качестве входной переменной для промежуточной модели выбран показатель ФС 50%.

Таблица 1. Корреляции показателей качества.

Показатель	ТНК	ФС 10%	ФС 50%	ФС 90%	ФС 95%	Вязкость
ТНК	1	0,7349	0,1196	-0,1018	-0,1102	0,1485
ФС 10%	0,7349	1	0,6274	0,3506	0,3048	0,6470
ФС 50%	0,1196	0,6274	1	0,8962	0,8556	0,9383
ФС 90%	-0,1018	0,3506	0,8962	1	0,9881	0,8447
ФС 95%	-0,1102	0,3048	0,8556	0,9881	1	0,8011
Вязкость	0,1485	0,6470	0,9383	0,8447	0,8011	1

Для построения промежуточной модели использовалась нейронная сеть, содержащая 1 скрытый слой, состоящий из 1 нейрона (гиперболический тангенс в качестве функции активации). Для обучения нейронной сети использован метод байесовской регуляризации. В качестве модели в составе виртуального анализатора использована линейная модель в силу ее широкой распространенности в производственных условиях.

Проведено деление имеющихся лабораторных данных по вязкости при 40°C на обучающую и тестовую выборку в пропорции 50/50. Для получения дополнительного

сегмента данных обучающей выборки использовано 201 значение показателя ФС 50% содержащееся во временном интервале, соответствующем обучающей выборке.

Для обучения ВА использованы следующие выборки:

- выборка 1, содержащая 15 наблюдений из лабораторных данных;
- выборка 2, содержащая 15 наблюдений из лабораторных данных и 201 наблюдение из полученного ДСД;
- выборка 3, содержащая 201 наблюдение из полученного ДСД обучающей выборки.

В качестве входных переменных виртуального анализатора используются: расход орошения К1 (ФС-01, м³/ч), расход ВЦО (ФС-02, м³/ч), расход фракции среднего дистиллята (ФС-03, м³/ч), расход сырья К1 (ФС-04, м³/ч), температура в зоне над вводом сырья (ТС-01, °С), температура из Р2 в К3 (ТС-02, °С), температура ВЦО (ТС-03, °С), давление низа К1 (Р-01, МПа).

Критериями точности ВА являлись коэффициент детерминации (R^2) и средняя абсолютная ошибка (САО):

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \text{ САО} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|,$$

где Y_i – лабораторное значение показателя, \hat{Y}_i – оценка показателя виртуальным анализатором, $\bar{Y} = 1/n \sum_{i=1}^n Y_i$ – среднее арифметическое лабораторных значений показателя, n – количество наблюдений в выборке.

В таблице 2 представлены результаты тестирования ВА, обученных с использованием выборок 1, 2 и 3, содержащих только значения лабораторного анализа вязкости при 40°С.

Таблица 2. Результаты тестирования ВА

№ Обучающей выборки	Обучающая выборка		Тестовая выборка	
	R^2	САО	R^2	САО
1	0,8534	0,0247	-0,2857	0,1686
2	0,4938	0,0409	0,4066	0,1049
3	0,4476	0,0472	0,3983	0,1063

Полученные результаты показывают, что использование для обучения выборки, содержащей значения лабораторного анализа и ДСД, позволяет избежать переобучения модели и снизить САО ВА на тестовой выборке на 37.8% в сравнении с моделью, обученной на малой исходной выборке.

4. Заключение

Предложенный алгоритм расширения малой обучающей выборки для построения ВА редкоизмеряемого показателя качества на основе данных более частоизмеряемых показателей качества выходного продукта позволил повысить точность ВА вязкости фракции среднего дистиллята и снизить на 37.8% значение САО относительно виртуального анализатора полученного без применения расширения обучающей выборки.

Список литературы

1. Napoli G., Xibilia M.G. Soft Sensor design for a Topping process in the case of small datasets // Computers and Chemical Engineering. 2011. Vol. 35, No. 11. P. 2447-2456.

2. Torgashov A., Samotylova S., Yang F. Soft sensors development for industrial reactive distillation processes under small training datasets // *Computer Aided Chemical Engineering*. 2022. Vol. 49. P. 937-942.
3. Li D, Lin L., Peng L. Improving learning accuracy by using synthetic samples for small datasets with non-linear attribute dependency // *Decision Support Systems*. 2014. Vol. 59. P. 286-295.
4. Chang C., Li D., Huang Y., Chen C. A novel gray forecasting model based on the box plot for small manufacturing data sets // *Applied Mathematics and Computation*. 2015. Vol. 265. P. 400-408.
5. Chen Z., Zhu B., He Y., Yu L. A PSO based virtual sample generation method for small sample sets: Applications to regression datasets // *Engineering Applications of Artificial Intelligence*. 2017. Vol. 59. P. 236-243.
6. He Y., Hua Q., Zhu Q., Lu S. Enhanced virtual sample generation based on manifold features: Applications to developing soft sensor using small data // *ISA Transactions*. 2022. Vol. 126. P. 398-406.
7. Zhang X., Xu Y., He Y., Zhu Q. Novel manifold learning based virtual sample generation for optimizing soft sensor with small data // *ISA Transactions*. 2021. Vol. 109. P. 229-241.
8. Zhu Q., Hou K., Chen Z., Gao Z., Xu Y., He Y. Novel virtual sample generation using conditional GAN for developing soft sensor with small data // *Engineering Applications of Artificial Intelligence*. 2021. Vol. 106. No. 104497.